MASTER

C00-2874-30

SIAM INSTITUTE FOR MATHEMATICS AND SOCIETY

33 SOUTH 17TH STREET • PHILADELPHIA, PA. 19103 • (215) LOCust 4-2930

BOARD OF DIRECTORS

DONALD L. THOMSEN, JR.
President

Reply To: SIMS
97 Parish Road South
New Canaan, Conn. 06840
(203) 966-1008

JOEL E. COHEN
Rockefeller University

WILLIAM T. GOLDEN
New York, New York

B. RICHARD HAND
Research Corporation

BROCKWAY McMILLAN
Bell Telephone Laboratories

WILLIAM F. MILLER
Stanford University

VAUN A. NEWILL
Exxon Corporation

EMANUEL R. PIORE
Rockefeller University

MINA S. REES
City University of New York

WERNER C. RHEINBOLDT
University of Maryland

DONALD L. THOMSEN, JR.
SIAM Institute for Mathematics
and Society

SIMS THREE YEAR STUDY ON STATISTICS AND ENVIRONMENTAL FACTORS IN HEALTH

Technical Progress Report No. 2

Table of Contents

Page

Introduction	1
Progress Report -- SIMS	2
Progress Report -- Columbia	20
Progress Report -- Stanford	33

NOTICE

This report was prepared as an account of work sponsored by the United States Government. Neither the United States nor the United States Department of Energy, nor any of their employees, nor any of their contractors, subcontractors, or their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness or usefulness of any information, apparatus, product or process disclosed, or represents that its use would not infringe privately owned rights.

December 1977

EB
DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

Prepared For
THE U.S. ENERGY RESEARCH AND DEVELOPMENT ADMINISTRATION
UNDER CONTRACT NO. EY-76-S-02-2874

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency Thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

Introduction

The three progress reports which follow for SIMS, Columbia, and Stanford should be reviewed by referencing where indicated the original "Proposal for SIMS Three Year Study on Statistics and Environmental Factors in Health" (February 1976). Future plans for SIMS, Columbia, and Stanford will also where appropriate make reference to this same basic February 1976 Proposal for the three year Study.

SIMS Progress Report for Second Year

Referencing the complete Proposal for SIMS Three Year Study on Statistics and Environmental Factors in Health (February 1976), SIMS activity continued to focus on those items discussed on pages 15-17, 88, and 89.

Review Panel. The Review Panel (see page 15 of February 1976 Proposal) held its second meeting in New York on August 20, 1977. Members* of the Panel present were Dr. William G. Cochran (Harvard), Dr. Vaun A. Newill (Exxon), Dr. Donald L. Thomsen, Jr. (SIMS, ex officio), and Dr. John R. Totter (Oak Ridge National Laboratory). By invitation Dr. Joseph L. Fleiss attended for Columbia and Dr. Paul Switzer for Stanford. Both Drs. Fleiss and Switzer presented summaries of their activities during which time the Panel Members asked questions and made suggestions. Minutes of the meeting were prepared by Dr. Thomsen and are included separately in Appendix A.

Communication (see page 88 of February 1976 Proposal) by members of the Study has continued to take place in the form of professional meetings and conference activity. At professional meetings during 1977, the following presentations were made:

Columbia

- (1) J. M. Cuzick
"Asymptotic normality of chi-square in $m \times n$ tables with n large and small cell expectations"
Spring Statistics Meetings,
Chapel Hill, N.C., 1977.
- (2) J. M. Cuzick
"The joint distribution of the maxima of two air pollution time series" (with Yash Mittal, Stanford University)
Biometric Society and IMS joint meeting,
Stanford, June 1977.

*Members of the Panel are: Dr. William G. Cochran (Harvard University), Dr. Mark Kac (Rockefeller University), Dr. Vaun A. Newill (Exxon Corporation), Dr. Donald L. Thomsen, Jr. (SIMS, ex officio), Dr. John R. Totter (Oak Ridge National Laboratory, DOE), and Dr. James L. Whittenberger (Harvard School of Public Health).

- (3) I. F. Goldstein and L. E. Rausch
"Time series analysis of records of asthmatic attacks in
New York City"
International Epidemiological Association,
San Juan, Puerto Rico, September 1977.

Stanford

- (1) Yash Mittal
"Use of averaging maxima in air pollution data"
Fourth Symposium on Statistics and the Environment,
Washington, May, 1976. (not listed in 1976 report)
- (2) Yash Mittal
"The joint distribution of the maxima of two air pollution time
series" (with J. M. Cuzick, Columbia University)
Biometric Society and IMS joint meeting,
Stanford, June, 1977.
- (3) Thomas W. Sager.
"A new approach to regression in certain time/space series
problems"
SIMS Research Application Conference on Time Series and Ecology
Alta, Utah
June 27-July 2, 1977

"Statistical issues in air pollution assessment"
Air Pollution Seminar, Department of Meteorology
California State University, San Jose
November 30, 1977
- (4) Paul Switzer
"Distribution of air pollutants in the San Francisco Bay Area"
Center for Geostatistics, Fontainebleau, France
September 20, 1977

"Estimation of spatial distributions from point sources
with application to air pollution measurement"
International Statistical Institute, Delhi, India
December 6, 1977
- (5) Alice S. Whittemore
"Mathematical models of carcinogenesis: epidemiological
implications"
University of California, Los Angeles
February, 1977

University of California, Riverside
March, 1977

California State University, Hayward
March, 1977

M. D. Anderson Hospital and Tumor Institute
Houston, Texas
December, 1977

A second Conference is planned for 1978 as per the February 1976 Proposal (page 88). The title of the second Conference is "Energy and Health"; co-chairmen of the Conference are Dr. Norman E. Breslow, University of Washington (Seattle) and Dr. Alice S. Whittemore, Hunter College, City University of New York, and Stanford University. (Dr. Whittemore is an active member of the SIMS Study at Stanford University, where she currently is visiting professor.) Additional members of the Conference Committee are Dr. Yvonne M. Bishop, Harvard University School of Public Health, and Dr. Donald L. Thomsen, Jr., SIMS. A proposal for the Conference has been prepared and submitted for joint funding by DOE and NSF.

A session entitled "Statistics and Environmental Factors in Health" has been proposed with the assistance of SIMS and accepted by the AAAS for the Annual Meeting in Washington D.C. in February 1978. The Session has been arranged and will be chaired by Dr. Whittemore.

Publications (see page 89 of February 1976 Proposal). A complete list (sequential) of publications from the Study with three subdivisions (1) publications for 1976, (2) publications for 1977, and (3) publications in preparation has been prepared and is presented in Appendix B. A second edition of the *News* of the Study is in preparation and is scheduled to be distributed by the end of 1977. An "order form" for all publications has been prepared; it is updated approximately every three months; the latest order form appears on page 5. The current distribution list is presented in Appendix C.

Please send to me at the address below the following reports and publications from the SIMS Study on Statistics and Environmental Factors in Health:

- A. ☐ Please send all publications listed in section D and all future publications.
 B. ☐ Please send titles only of future publications.
 C. ☐ Please send all publications listed in section D.
 D. ☐ Please send only those reports checked below.

- ___ RELATING SPATIAL DISTRIBUTIONS OF POLLUTANTS TO HEALTH EFFECTS,* by Thomas W. Sager (Stanford), July 15, 1976
 ___ ROBUSTNESS TO MISSPECIFICATIONS OF CORRELATION MODELS FOR THE PROBLEM OF INTERPRETING AIR POLLUTION MAPS,* by Ruth Sheshinski (Stanford), July 22, 1976
 ___ MAXIMA OF PARTIAL SAMPLES IN GAUSSIAN SEQUENCES,* by Yashaswini Mittal (Stanford), Aug. 1, 1976
 ___ ROBUSTNESS OF LINEAR ESTIMATORS TO MISSPECIFICATION OF TREND,* by Ruth Sheshinski (Stanford), Aug. 16, 1976
 ___ STUDY OF THE MAXIMA OF PARTIAL SAMPLES OF OXIDANT CONCENTRATIONS IN THE BAY AREA,** by Yashaswini Mittal (Stanford), Sept. 1, 1976
 ___ AN ANALYSIS OF AIR POLLUTION DATA IN KAWASAKI CITY, JAPAN,** by Yasutoshi Washio (Stanford), Sept. 17, 1976
 ___ CORRELATION IN CONDITIONAL DISTRIBUTIONS: SOME SUGGESTIONS FOR MEASURING THE RELATIONSHIPS BETWEEN TWO SETS OF POLLUTANTS WHEN THE EFFECT OF WEATHER HAS BEEN REMOVED,** by June Juritz (Stanford), Oct. 11, 1976
 ___ THE USE OF DOUBLY STOCHASTIC POISSON PROCESSES IN ESTIMATING HEALTH EFFECTS DUE TO AIR POLLUTION, by Jack M. Cuzick (Columbia), July 1976
 ___ SHELF LIST OF REPORTS ON AIR POLLUTION AND HEALTH FILED IN THE GIRSHICK LIBRARY, STANFORD UNIVERSITY,** by Bonnie Hole (Stanford), Oct. 25, 1976
 ___ THE INTERPOLATION OF AIR POLLUTION: A STOCHASTIC APPROACH,** by Ray Faith (Stanford), Nov. 1, 1976
 ___ TIGHT UPPER AND LOWER BOUNDS FOR CORRELATION OF BIVARIATE DISTRIBUTIONS ARISING IN AIR POLLUTION MODELING,* by Dick DeVeaux (Stanford), Nov. 11, 1976
 ___ SCIENTIFIC BASE FOR SULFUR OXIDE AIR POLLUTION STANDARDS, by Inge F. Goldstein (Columbia), Nov. 19, 1976
 ___ NEWS:- SIMS Study on Statistics and Environmental Factors in Health, Vol. 1, No. 1, October 1976
 ___ TECHNICAL PROGRESS REPORT NO. 1: SIMS THREE YEAR STUDY ON STATISTICS AND ENVIRONMENTAL FACTORS IN HEALTH, December 1976 (revised July 1977)
 ___ A NEW MEASURE OF ASSOCIATION FOR SPATIAL VARIABLES,* by Dag Tjøstheim (Stanford), March 17, 1977
 ___ MULTIPLE THREE-DECISION RULES FOR 2^k -FACTORIAL SIMPLE EFFECTS,* by R. Bohrer, W. Chow, R. Faith, V. M. Joshi, C. F. Wu (Stanford), March 22, 1977
 ___ ASYMPTOTIC NORMALITY OF χ^2 IN $m \times n$ TABLES WITH n LARGE AND SMALL CELL EXPECTATIONS, by Jack Cuzick (Columbia), June 1977
 ___ THE RELATION OF AIR POLLUTION TO MORTALITY IN NEW YORK CITY, by Inge F. Goldstein (Columbia), Martin Goldstein, and Leon Landovitz, June 1977 (Revised December 1977)
 ___ ESTIMATION OF SPATIAL DISTRIBUTIONS FROM POINT SOURCES WITH APPLICATION TO AIR POLLUTION MEASUREMENT,* by Paul Switzer (Stanford), July 1, 1977
 ___ ON EXTREME VALUES OF SAMPLED AND CONTINUOUS STOCHASTIC DATA,* by M. R. Leadbetter (Stanford), Aug. 1, 1977
 ___ TIME SERIES ANALYSIS OF MORBIDITY DATA FOR ASSESSMENT OF ACUTE ENVIRONMENTAL HEALTH EFFECTS, by Inge F. Goldstein and Lawrence E. Rausch (Columbia), October, 1977
 ___ A NEW APPROACH TO REGRESSION IN CERTAIN TIME/SPACE SERIES PROBLEMS,* by Thomas W. Sager (Stanford), Oct. 14, 1977
 ___ APPLICATION OF SOME ROBUST STATISTICAL PROCEDURES TO AIR POLLUTION DATA IN THE SAN FRANCISCO BAY AREA,** by Chen-Hui Foo (Stanford), September 1977
 ___ COLLAPSIBILITY OF MULTIDIMENSIONAL CONTINGENCY TABLES TO TABLES OF SMALLER DIMENSION,* by Alice S. Whittemore (Stanford), Dec. 6, 1977
 ___ THE AGE DISTRIBUTION OF HUMAN CANCER FOR CARCINOGENIC EXPOSURES OF VARYING INTENSITY,* by Alice S. Whittemore (Stanford), Dec. 9, 1977
 ___ EXPOSURE DATA FOR THE STUDY OF ACUTE HEALTH EFFECTS OF AIR POLLUTION, by Inge F. Goldstein (Columbia), Dec. 1977
 ___ METHODOLOGICAL PROBLEMS ARISING FROM THE CHOICE OF AN INDEPENDENT VARIABLE IN LINEAR REGRESSION, WITH APPLICATION TO AN AIR POLLUTION EPIDEMIOLOGICAL STUDY, by Inge F. Goldstein, Martin Goldstein, Joseph L. Fleiss, and Leon Landovitz (Columbia), Dec. 1977
 ___ TECHNICAL PROGRESS REPORT NO. 2: SIMS THREE YEAR STUDY ON STATISTICS AND ENVIRONMENTAL FACTORS IN HEALTH, December 1977.

December 1977

*Technical Report **Working Paper

Signed _____

Name (PLEASE PRINT) _____

Address _____

City _____

State _____

ZIP _____

Comprehensive Dissemination of Study's Results. SIMS in the Proposal (February 1976) on page 89, last paragraph, stated: "Publication plans include the preparation at the end of the three year period of one or more volumes incorporating results of the Study; these volumes it is expected will be in a form useful to investigators and practitioners in the field of environmental health. Both the Columbia and Stanford Centers plan to assist in the preparation of such a volume."

These plans were raised specifically and discussed in the Review Panel Meeting of August 20, 1977 (see Minutes, Appendix A). From other sources SIMS has received comment that to ensure that results of the Study are made available, several forms should be considered, specifically two of which are:

- (1) A special version of the Study prepared for the most part by someone who has the ability to both understand the technical results and write about them in an orderly and understandable way.
- (2) A collection of the papers and reports in a single volume (or more if necessary) with suitable groupings into sections and with short introductions to each section.

The second possibility above could be accomplished by members of the Study, with some of whom discussions in this connection have already been held; the first possibility would most likely mean engaging a special person to do the work. Decisions and plans with respect to both these possibilities (and others which may be proposed) will be made during the upcoming third year, and the actual work would most likely be set up and carried out the first part of 1979.

This effort is regarded as one of the most important activities of the Study, particularly so in view of the number of reports and publications which have appeared and which are in progress. From the Budget point of view, this is reflected in the item Publications (Reference February 1976 Proposal, page 93.).

Feedback: New Directions for the Study

Currently SIMS is exploring informally possibilities in connection with the six-city pollution health study (Watertown, MA, Kingston-Harriman, TN, St. Louis, MO, Steubenville, OH, Portage, WI, and Topeka, KS) under the direction of Drs. Bishop, Ferris, and Speizer at the Harvard School of Public Health.* Several informal conversations have been held with one member of the Harvard group. Considerable quantities of data are being collected. It is anticipated that some of the participants in the SIMS Study at Columbia and Stanford may be involved with respect to analyzing that data from a methodological point of view. In this connection, one of the technical reports from the SIMS Study at Stanford (Switzer: "Estimation of Spatial Distributions from Point Sources with Application to Air Pollution Measurement") has been cited by a member of the Harvard group as being potentially useful; steps are being actively taken by SIMS to determine the practicality of this application.

Also, the Bay Area Pollution Control District (BAAPCD), Stanford University, and SIMS are actively exploring research collaboration for the future beyond the three years of the current Study. In this connection, the BAAPCD would become in effect an active "laboratory" for those at Stanford participating in the next phase of this Study.

*Yvonne M. M. Bishop, B. G. Ferris, Jr. and F. E. Speizer, "Problems in Relating Health Effects to Air Pollution," *Environmental Health: Quantitative Methods* (Proceedings of SIMS Research Application Conference on Environmental Health held July 1976), ed. Alice S. Whittemore (Philadelphia, 1976), pp. 215-228.

With respect to Funding (see pages 16, 17 of February 1976 Proposal), the Budget for the Three Year Study has been set at \$838,900. in total costs (direct and indirect). Approvals have been received for the first two years in an amount totalling \$558,700. from three government agencies, EPA, ERDA (now DOE), and NSF, and two private foundations (Rockefeller and Sloan). Details on the exact amounts are presented in Appendix D of this section.

SIMS THREE YEAR STUDY ON STATISTICS AND ENVIRONMENTAL FACTORS IN HEALTH

Review Panel Meeting - August 20, 1977

The following Panel members were in attendance:

Dr. W. G. Cochran, Harvard U.
Dr. V. A. Newill, Exxon Corp.
Dr. J. R. Totter, Oak Ridge National Laboratory (ERDA)
Dr. D. L. Thomsen, Jr., SIMS (ex officio)

The following guests were in attendance:

Dr. J. L. Fleiss, Columbia U.
Dr. P. Switzer, Stanford U.

Regarding activity at Columbia in the SIMS Study, Dr. Fleiss reported on work related to the New York City Aerometric Network and noted that two reports would be forthcoming mid or end of fall '77. Results to date indicate that useful air pollution information is derived from contiguous measuring stations for local areas. However, useful information cannot be obtained from a single station for the city as a whole. Dr. Fleiss further noted that results of analyzing emergency room data with respect to asthma produce health predictions, for example in regard to flu epidemics several days later. The work at Columbia is giving results which indicate what can and what cannot be deduced from the exposure data with respect to health effects.

As for new directions, Dr. Fleiss noted possible cooperation with the Environmental Health Sciences Center under Dr. Bernard Weinstein at the School of Public Health and work which would be directed along occupational lines in New Jersey industry. Dr. Newill expressed an interest in which particular industry might be involved -- such an industrial association being desirable and necessary to carry out this kind of work.

Dr. Fleiss reviewed the activities of two graduate students in the Study one of whom had been carrying out analysis of emergency room data.

Dr. Switzer reviewed progress on the part of the Stanford University Statistics Department in creating a center for the application of mathematical statistics to problems in air pollution. He noted that interest was obtained in holding seminars on air pollution effects on health. Attendance was formed by graduate students, faculty and outsiders, some from as far away as Sacramento and Los Angeles. Initially the talks were of a general nature by outsiders -- now they are more often presented by their own faculty and graduate students -- this reflecting the learning process which has taken place.

Next Dr. Switzer detailed the work of the eight graduate students in the Study. Topics included measuring variations of pollution in relation to

-2-

variations of weather, discrete maxima versus continuous maxima, applications of robust regression techniques to weather effect on air pollution, low dose extrapolations, fitting models to contingency tables, estimation of correlations from transformed variables, and interpolating between air pollution measurement stations.

In addition, Dr. Switzer, by way of example, briefly outlined selected research efforts of the faculty. Research has focused on population and health effects re pollution contours, the quality of air monitoring data for particulates which gave rise to the question of maxima with partial samples (i.e. relation of a subsequence of a time series to maxima), and behavior of joint distribution of maxima over time as applied to same pollutant at two different stations.

Finally, Dr. Switzer noted the Library is growing and a new shelf list would soon be ready.

Future possibilities were reviewed by the group. Dr. Fleiss noted one course at Columbia (already indicated above) in connection with work in occupational safety and health in New Jersey at the same time as maintaining contacts with New York City. Dr. Fleiss reported one related development in that shortly the Biostatistics Division will have a Ph.D. Program.

At Stanford Dr. Switzer told about the positive feeling on the part of the faculty -- that everyone was pleased with the program and that the whole thing had been beneficial. The feeling was that the effort should continue.

By way of background, Dr. Switzer pointed out that a group at Stanford in biostatistics had in the Stanford Medical School a kind of laboratory; a comparable facility for the group working on air pollution and health would be desirable; in this connection the Bay Area Air Pollution Control District (BAAPCD) has emerged as a facility comparable to that of a laboratory and some more formal ties would be desirable in the future, particularly if the effort is to continue after the three years. Details as to space, data access, equipment, and budgets were then discussed.

During the course of the two and one half hour meeting Panel members asked a number of searching questions regarding the results to date and future directions. They made many comments and offered much useful advice. Note was taken of the paucity of good epidemiological data and qualified medical epidemiologists. Relation to other programs and studies was discussed, particularly the six-city study at Harvard under the direction of Bishop, Ferris, and Speizer.

Selected comments in summary form by individual Panel Members follow:

Dr. Cochran: Noted wide applicability of the research on partial samples; raised question of health data in Study; questioned form and scope of the final report; summarized the general interest in the Study with respect to continuing after three years.

Dr. Newill: Voiced the potential contribution of the Study to the regulatory issue -- those in the regulatory business do not understand "variance" -- do not admit it exists -- much national resource spent without good underpinning. Suggested Speizer give a lecture at Stanford.

Dr. Totter: Noted value of finding out what is not going to work (ref. to Columbia effort); recalled his impatience to see health effects brought in, but now less concerned in view of CHESS having been too rapidly set up and after having become recently more appreciative of many of the problems involved.

Panel members held a brief session after lunch to review some of the points raised during the morning session.

These minutes were prepared from a complete recording of the meeting.

D. L. Thomsen, Jr.

November 28, 1977

LIST OF PUBLICATIONS

1976

ERDA NO.

- COO-2874-1 RELATING SPATIAL DISTRIBUTIONS OF POLLUTANTS TO HEALTH EFFECTS,*
by Thomas W. Sager (Stanford), July 15, 1976
- COO-2874-2 ROBUSTNESS TO MISSPECIFICATIONS OF CORRELATION MODELS FOR THE
PROBLEM OF INTERPRETING AIR POLLUTION MAPS,* by Ruth Shenshinski
(Stanford), July 22, 1976
- COO-2874-3 MAXIMA OF PARTIAL SAMPLES IN GAUSSIAN SEQUENCES*;
by Yashaswini Mittal (Stanford), August 1, 1976
- COO-2874-4 ROBUSTNESS OF LINEAR ESTIMATORS TO MISSPECIFICATION OF TREND*;
by Ruth Sheshinski (Stanford), August 16, 1976
- COO-2874-5 STUDY OF THE MAXIMA OF PARTIAL SAMPLES OF OXIDANT CONCENTRATIONS
IN THE BAY AREA,** by Yashaswini Mittal (Stanford), September 1, 1976
- COO-2874-6 AN ANALYSIS OF AIR POLLUTION DATA IN KAWASAKI CITY, JAPAN,**
by Yasutoshi Washio (Stanford), September 17, 1976.
- COO-2874-7 CORRELATION IN CONDITIONAL DISTRIBUTIONS: SOME SUGGESTIONS
FOR MEASURING THE RELATIONSHIPS BETWEEN TWO SETS OF POLLUTANTS
WHEN THE EFFECT OF WEATHER HAS BEEN REMOVED,** by June Juritz
(Stanford), October 11, 1976
- COO-2874-8 THE USE OF DOUBLY STOCHASTIC POISSON PROCESSES IN ESTIMATING
HEALTH EFFECTS DUE TO AIR POLLUTION, by Jack M. Cuzick (Columbia),
July 1976.
- COO-2874-9 THE RELATION OF AIR POLLUTION TO MORTALITY: A CRITIQUE,
by Inge F. Goldstein, Martin Goldstein, and Leon Landovitz
(Columbia), 1976
- COO-2874-10 SHELF LIST OF REPORTS ON AIR POLLUTION AND HEALTH FILED IN
THE GIRSHICK LIBRARY, STANFORD UNIVERSITY,** by Bonnie Hole
(Stanford), October 25, 1976
- COO-2874-11 THE INTERPOLATION OF AIR POLLUTION: A STOCHASTIC APPROACH,**
by Ray Faith (Stanford), November 1, 1976
- COO-2874-12 TIGHT UPPER AND LOWER BOUNDS FOR CORRELATION OF BIVARIATE
DISTRIBUTIONS ARISING IN AIR POLLUTION MODELING,* by Dick
DeVeaux (Stanford), November 11, 1976
- COO-2874-13 SCIENTIFIC BASE FOR SULFUR OXIDE AIR POLLUTION STANDARDS,
by Inge F. Goldstein (Columbia), November 19, 1976
- COO-2874-14 NEWS:- SIMS STUDY ON STATISTICS AND ENVIRONMENTAL FACTORS
IN HEALTH, Volume 1, No. 1, October 1976
- COO-2874-15 TECHNICAL PROGRESS REPORT: SIMS THREE YEAR STUDY
ON STATISTICS AND ENVIRONMENTAL FACTORS IN HEALTH,
December 1976

*Technical Report

**Working Paper

LIST OF PUBLICATIONSERDA NO.

1977

- COO-2874-16 A NEW MEASURE OF ASSOCIATION FOR SPATIAL VARIABLES,*
by Dag Tjøstheim (Stanford), March 17, 1977
- COO-2874-17 MULTIPLE THREE-DECISION RULES FOR 2^k -FACTORIAL SIMPLE EFFECTS,*
by R. Bohrer, W. Chow, R. Faith, V. M. Joshi, C. F. Wu (Stanford),
March 22, 1977
- COO-2874-18 ASYMPTOTIC NORMALITY OF χ^2 IN $m \times n$ TABLES WITH n LARGE AND
SMALL CELL EXPECTATIONS, by Jack Cuzick (Columbia), June 1977
- COO-2874-19 THE RELATION OF AIR POLLUTION TO MORTALITY IN NEW YORK CITY,
by Inge F. Goldstein (Columbia), Martin Goldstein, and Leon
Landovitz, June 1977 (Revision of COO-2874-9)
- COO-2874-20 ESTIMATION OF SPATIAL DISTRIBUTIONS FROM POINT SOURCES WITH
APPLICATION TO AIR POLLUTION MEASUREMENT,* by Paul Switzer
(Stanford), July 1, 1977
- COO-2874-21 ON EXTREME VALUES OF SAMPLED AND CONTINUOUS STOCHASTIC DATA,*
by M. R. Leadbetter (Stanford), August 1, 1977
- COO-2874-22 TIME SERIES ANALYSIS OF RECORDS OF ASTHMATIC ATTACKS
IN NEW YORK CITY,** by Inge F. Goldstein and
Lawrence E. Rausch (Columbia), September 1977
- COO-2874-23 A NEW APPROACH TO REGRESSION IN CERTAIN TIME/SPACE
SERIES PROBLEMS,* by Thomas W. Sager (Stanford),
October 14, 1977
- COO-2874-24 APPLICATION OF SOME ROBUST STATISTICAL PROCEDURES
TO AIR POLLUTION DATA IN THE SAN FRANCISCO BAY AREA,**
by Chen-Hui Foo, September 1977
- COO-2874-25 COLLAPSIBILITY OF MULTIDIMENSIONAL CONTINGENCY TABLES TO TABLES
OF SMALLER DIMENSION,* by Alice S. Whittemore (Stanford),
December 6, 1977
- COO-2874-26 THE AGE DISTRIBUTION OF HUMAN CANCER FOR CARCINOGENIC EXPOSURES
OF VARYING INTENSITY,* by Alice S. Whittemore (Stanford),
December 9, 1977
- COO-2874-27 EXPOSURE DATA FOR THE STUDY OF ACUTE HEALTH EFFECTS OF AIR
POLLUTION, by Inge F. Goldstein (Columbia), December 1977

(List continued on next page)

*Technical Report

**Working Paper

LIST OF PUBLICATIONS

1977 (continued)

ERDA NO.

- COO-2874-28 METHODOLOGICAL PROBLEMS ARISING FROM THE CHOICE OF
AN INDEPENDENT VARIABLE IN LINEAR REGRESSION, WITH
APPLICATION TO AN AIR POLLUTION EPIDEMIOLOGICAL STUDY,
by Inge F. Goldstein, Martin Goldstein, Joseph L. Fleiss,
and Leon Landovitz (Columbia), December 1977
- COO-2874-29 THE RELATION OF AIR POLLUTION TO MORTALITY, by
Inge F. Goldstein, Martin Goldstein, and
Leon Landovitz (Columbia), December 1977
(Revision of COO-2874-19)
- COO-2874-30 TECHNICAL PROGRESS REPORT NO. 2: SIMS THREE YEAR STUDY
ON STATISTICS AND ENVIRONMENTAL FACTORS IN HEALTH,
December 1977

PUBLICATIONS IN PREPARATION

December 1977

SIMS

NEWS:- SIMS STUDY ON STATISTICS AND ENVIRONMENTAL FACTORS IN HEALTH,
Vol. 1, No. 2.

Columbia

Goldstein, I. F., Landovitz, L., and Fleiss, J. L. "Air pollution patterns in New York City, III." Spacial distributions of air pollutants. Paper in preparation.

Goldstein, I. F., Fleiss, J. L., and Landovitz, L. "Air pollution patterns in New York City, IV." Principal components analyses. Paper in preparation.

Stanford

Tjøstheim, Dag. "On the problem of estimating the autocorrelation function of spatio-temporal variables." Technical report in process.

Chernick, Michael. "A limit theorem for the maximum of an exponential autoregressive process." Technical report in process.

Homer, Jack B. "Techniques for discriminating between normal and lognormal populations with special application to air pollution data." Working paper in process.

Distribution List for SEFH Reports - December 1977

Norman Agin
Mathtech Inc.
P.O. Box 2392
Princeton, NJ 08540

B. Robert Agins
4075 Berkshire E
Deerfield Beach, FL 33441

D. S. Ballantine
Div. of Biomed. & Env. Research
U.S. ERDA
Washington, DC 20545

Robert W. Beadle
AES-BER
Washington, DC 20545

Yvonne M. Bishop
Dept. of Biostatistics
Harvard School of Public Health
677 Huntington Ave.
Boston, MA 02115

William G. Cochran
P.O. Box 249
South Orleans, MA 02662

Joel E. Cohen
Rockefeller University
1230 York Ave.
New York, NY 10021

B. H. Colvin
National Bureau of Standards
A-438, Admin. Bldg.
Washington, DC 20234

John Paul Creason
SDMO, HERL
U.S. EPA (MD-55)
Research Triangle Park, NC 27711

Jack M. Cuzick
Dept. of Mathematical Statistics
Columbia University
New York, NY 10027

James Q. Denton
Math Sciences Section
NSF
1800 G St. NW
Washington, DC 20550

Joseph L. Fleiss
Division of Biostatistics
Columbia University, Sch. of Public Health
600 West 168th St.
New York, NY 10032

F. P. Glick
Dept. of Health Care & Epidemiology
University of British Columbia
Vancouver, B.C. VGT-1W5

John R. Goldsmith
Epidemiological Studies Laboratory
California State Department of Health
2151 Berkeley Way
Berkeley, CA 94704

Inge F. Goldstein
Columbia University, Sch. of Public Health
600 West 168th St.
New York, NY 10032

David G. Hoel
National Institute of Environmental
Health Sciences
P.O. Box 12233
Research Triangle Park, NC 27709

Mark Kac
Rockefeller University
1230 York Ave.
New York, NY 10021

James D. Koerner
Alfred P. Sloan Foundation
630 Fifth Ave.
New York, NY 10020

Simon Levin
Department of Ecology
Cornell University
Ithaca, NY 14850

Sidney Marks
Battelle Pacific Northwest Laboratory
Battelle Blvd.
Richland, WA 99352

Brockway McMillan
Bell Telephone Laboratories
Whippany Rd.
Whippany, NJ 07981

(continued)

Distribution List for SEFH Reports - December 1977
(continued)

S. C. Morris
Brookhaven National Laboratory
Building 475
Upton, NY 11973

William C. Nelson
Chief, Statistics and Data
Management Office
Health Effects Research Lab.
U.S. EPA
Research Triangle Park, NC 27711

Vaun A. Newill
Exxon Corporation
P.O. Box 45
Linden, NJ 07036

Paul Nutkowitz
Senior Environmental Planner
Systems Evaluation Unit
Tri-State Regional Planning Com.
One World Trade Center, 82 Floor
New York, NY 10048

Ingram Olkin
Department of Statistics
Stanford University
Stanford, CA 94305

Clifford H. Patrick
Group Leader, Health Effects
& Epidemiology
ORNL, Bldg. 4500-S, P.O. Box X
Oak Ridge, TN 37830

William H. Pell
Head, Mathematical Sciences Section
NSF
Washington, DC 20550

Mary Ann Potts
Dept. Health Care & Epidemiology
University of British Columbia
Vancouver, B.C.

Fred S. Roberts
Department of Mathematics
Rutgers University
New Brunswick, NJ 08903

Lewis H. Robinson
Bay Area Air Pollution Control Dist.
939 Ellis St.
San Francisco, CA 94109

Andrea Rosen
Regional Health Planner
North San Joaquin Valley
Health Systems Agency
3460 Oakdale Rd., Suite E
Oakdale Sylvan Plaza
Modesto, CA 95355

Ethel Salonen
Environmental Information Service
Main Library
SUNY at Stony Brook
Stony Brook, NY 11794

Patricia M. Scanlon
Pennie & Edmonds
330 Madison Ave.
New York, NY 10017

Richard G. Segers
Exxon Corporation
P.O. Box 153
Florham Park, NJ 07932

Steve Selvin
110 Haviland Hall, Sch. of Public Health
University of California, Berkeley
Berkeley, CA 94720

Elissa Speizman - 6 copies
U.S. Department of Energy
Contracts Management Office
Chicago Operations Office
9800 South Cass Ave.
Argonne, IL 60439

Paul Switzer
Department of Statistics
Stanford University
Stanford, CA 94305

(continued)

Distribution List for SEFH Reports - December 1977
(continued)

Gary H. Toenniessen
The Rockefeller Foundation
1133 Avenue of the Americas
New York, NY 10036

John R. Totter
Oak Ridge National Laboratory
P.O. Box Y
Oak Ridge, TN 37830

B. Vincent Viscomi
Dept. of Civil Engineering
Lafayette College
Easton, PA 18042

James L. Whittenberger
Harvard University
School of Public Health
665 Huntington Ave.
Boston, MA 02115

R. J. Fritz
P.O. Box 101
Florham Park, NJ 07932

SIMS - Appendix DTotal Funding - Status

<u>Year 1</u>	RF ¹	\$ 75,000.	)	
	ERDA ²	83,000.	)	
	NSF ³	22,400.	)	Received
	Sloan	30,000.	)	
	EPA ⁴	<u>45,000.</u>	)	
				\$255,400.
 <u>Year 2</u>	RF	\$ 65,000.		Received
	ERDA	83,000.		Received
	NSF	25,300.		Received
	Sloan	30,000.		Received
	EPA	<u>100,000.</u>		Approved
				303,300.
 <u>Year 3</u>	RF	\$ 60,000.		Being requested through renewal
	ERDA	84,300.		" " " "
	NSF	27,900.		" " " "
	Sloan	28,000.		" " " "
	EPA	<u>80,000.</u>		" " " "
				280,200.
				\$838,900.

December 1977

¹Rockefeller Foundation
²Energy Research and Development Administration
³National Science Foundation
⁴Environmental Protection Agency

Columbia University

1977 Progress Report and Plans for 1978

The Columbia group includes members of the Divisions of Biostatistics and Epidemiology of the School of Public Health and of the Department of Mathematical Statistics of the Graduate School of Arts and Sciences. Its objective is to develop and test statistical methods and models for describing the effects of the environment -- measured by pollution and weather variables -- on health -- so far measured only by respiratory diseases.

Research has continued on both the measurement of the extent of air pollution to which populations are exposed and the elucidation of the possible acute health effects of air pollution. The data analyzed were originally collected for a study of air pollution and emergency room visits (NIEHS grant 00899).

Those participating at Columbia University are:

School of Public Health

Joseph L. Fleiss (Biostatistics)
Inge F. Goldstein (Epidemiology)
Sylvan Wallenstein (Biostatistics)

Graduate Research Assistants

Neng-rong Lee (Cuzick)*
Lawrence Rausch (Goldstein)
Wen-jen Yeh (Cuzick)

Department of Mathematical Statistics

Jack M. Cuzick

Consultant

Leon F. Landovitz (City University of New York, Graduate Center)

*()'s indicate participant in SIMS project under whom graduate research assistant is or had been working.

I. The Measurement of Air Pollution (see Section III of original proposal, 2/76)

(1) Data analysis by Drs. Goldstein, Landovitz and Fleiss is just about complete on the daily recordings over a three year period of sulfur dioxide and suspended particulates (smokeshade) from the New York City aerometric network. We have concluded that these data are not suitable for epidemiological studies of acute health effects. This conclusion applies, strictly speaking, to the data from the New York City aerometric network; its applicability to other air monitoring networks remains to be ascertained.

Routinely collected aerometric data may, however, be suitable for the epidemiological study of long-term health effects. The reason for this assertion, simply stated, is that no physically reasonable patterns of daily or weekly air pollution variation have been found after intensive analysis of the data, but sensible patterns have emerged from analyses of longer-range (three month) averages. For example, we have found that the mean concentration of sulfur dioxide at a monitoring station is, approximately, a linearly increasing function of the station's height above street level (but not of its height above sea level). This finding is physically reasonable because sulfur dioxide is a diffusing gas whose sources are generally situated at heights above those of the monitoring stations.

In addition, provided we restrict attention to stations within the city and not on its periphery, we have found that stations close together have more similar mean levels of sulfur dioxide than stations far apart. What makes this a result worth stating is not its surprise value, given our physical knowledge of the environment, but its contrast to our previous inability to find any association at all between the distances separating the stations and their similarity in daily or weekly levels of sulfur dioxide.

As before, however, we have been unable to elucidate sensible patterns for smokeshade. A paper describing this research is being prepared for submission to Atmospheric Environment (Goldstein, Landovitz, Fleiss: "Air pollution patterns in New York City, III. Spatial distributions of air pollutants.").

(2) In spite of the compelling evidence that available data are inadequate for studies of acute health effects of air pollution, such studies continue. Drs. Goldstein, Fleiss and Landovitz have completed a methodological critique of one of the more important recent papers on this topic, that by Schimmel and Murawski ("The relation of air pollution to mortality". J. occup. Med., 1976, 18, 332). Using data on mortality from a variety of causes for New York City as a whole, but data on air pollution from only one of the 40 aerometric stations in the city (the laboratory in Harlem), they concluded from a series of regression analyses that sulfur dioxide is not implicated in excess daily mortality.

The methodological error we believe they made is a common one in regression analysis. It is often necessary, for reasons of economy or availability of data, to use as an independent variable not the "ideal" variable demanded by the hypothesis under study but some more easily accessible "practical" variable which is believed to be an adequate approximation to the ideal one. When the correlation between the practical and ideal variables is known, a mistake has been to use it in order to arrive at a unique "corrected" value for the regression coefficient of the dependent on the ideal independent variable. In actuality, one can never hope to find the desired regression coefficient in this way. The best one can do is obtain an interval, possibly quite wide, within which the desired coefficient lies.

Let X denote the ideal independent variable (average city-wide levels of sulfur dioxide, in the problem being considered), Y the practical independent variable (levels of sulfur dioxide measured at a single station), and Z the dependent variable (city-wide mortality). Let σ_x^2 , σ_y^2 and σ_z^2 denote the respective variances, ρ_{xy} and ρ_{yz} the correlations between the indicated variables, and $\beta_{z \cdot x}$ and $\beta_{z \cdot y}$ the regression coefficients of z on the ideal and practical variables, respectively.

Given the values of σ_x^2 , σ_y^2 , ρ_{xy} and $\beta_{z \cdot y}$ Schimmel and Murawski in effect took

$$\beta_{z \cdot x}^* = \frac{\sigma_y \rho_{xy}}{\sigma_x} \beta_{z \cdot y}$$

as the uniquely determined value for the desired regression coefficient. Actually, all one can validly assert is that the desired regression coefficient, $\beta_{z \cdot x}$, is within the interval

$$\beta_{z \cdot x}^* \pm \frac{\sigma_z}{\sigma_x} (1 - \rho_{xy}^2)^{1/2} (1 - \rho_{yz}^2)^{1/2}.$$

The widths of the resulting intervals were between 15 and 150 times the values of the "corrected" coefficients themselves, raising serious doubts about the correctness of Schimmel and Murawski's conclusions. A paper describing this analysis is being submitted for publication in the American Journal of Epidemiology (Goldstein, Goldstein, Fleiss and Landovitz: "Methodological problems arising from the choice of an independent variable in linear regression.").

(3) The final research endeavor being completed relevant to the measurement of air pollution has involved principal components analysis. Such analyses have been applied by Drs. Goldstein, Fleiss and Landovitz to data from all 40 stations comprising the New York City aerometric network,

as well as to data from the stations in and around Harlem and Bedford-Stuyvesant, the two study areas which contain the populations yielding our health data (see below).

With respect to the city-wide data, the results indicated that pollution patterns monitored by the 40 aerometric stations were so dissimilar that the first principal component could not be trusted to characterize the entire city's pollution, and that the other components made little physical (geographic or meteorological) sense. With respect to the data from Harlem and Bedford-Stuyvesant separately, the first principal components tended both to account for appreciable proportions of the variance (60% or more) and to be measurable as simple unweighted averages of the readings from the several stations involved. However, no clear-cut separations were found between the readings from the two sets of stations. A paper describing this work is being prepared for submission to Atmospheric Environment (Goldstein, Fleiss and Landovitz: "Air pollution patterns in New York City, IV. Principal components analyses.").

II. The Measurement of Health Effects (see section II of original proposal)

(4) Sharply focused analyses are still under way of the daily visits for asthma and other respiratory conditions to the emergency rooms of the hospitals serving the two inner city areas we have decided to concentrate on, Harlem and Bedford-Stuyvesant. One of the strategies adopted by Drs. Goldstein and Wallenstein may be described simply: can we identify specific periods of time in which unusually large numbers of people showed up at the emergency rooms for particular complaints? If the existence of

such periods could be established, and if the periods coincided in the two areas of the city we have studied, then environmental factors present city-wide on those days would be implicated and a search for them would be indicated.

Generally, asthma visits to the study hospitals were found to be well-fit by Poisson distributions in each of the seasons studied. All three fall seasons (from 1969 to 1971), however, had several days with high numbers of visits, and these days were usually the same at all of the emergency rooms. These simultaneous clusterings of large numbers of visits make them especially unlikely events.

This strategy and these results were presented at the September, 1977 meeting of the International Epidemiology Association in San Juan. Further research is being conducted by Dr. Wallenstein into possibly more appropriate means for characterizing days with unusually many visits and for detecting clusters of such days. Concurrently, the environmental conditions present during the clearly unusual days are being investigated.

(5) The strategy just described is useful for implicating city-wide environmental factors in emergency room visits for respiratory disorders. A different strategy is called for when the aim is to identify localized environmental conditions. A simple statistical procedure is to calculate chi square-like statistics for homogeneity to compare the distributions of daily emergency room visits at the study hospitals, and to identify for further study those days which make the greatest contributions to the total chi square.

An important theoretical result has been found concerning the asymptotic distribution of this statistic. Consider the case of an $m \times n$ cross-classification table with m fixed (m would correspond to the number of

hospitals studied) and with the total number of observations proportional to n where n is large (n would correspond to the number of days studied). The exact mean and variance of the chi square statistic have been known for nearly 40 years, but Dr. Cuzick has succeeded in proving that, asymptotically as n tends to infinity at the same rate as the number of observations, the statistic is normally distributed. This result and some of its applications to our data were presented at the spring statistics meetings in Chapel Hill.

Results show that larger chi square values occur when considering upper respiratory disorders rather than asthma or lower respiratory disorders. Thus, some upper respiratory illnesses may be caused by geographically localized environmental conditions. In addition, the chi square values for all three kinds of conditions tend to be smaller when homogeneity is examined within geographic areas than when the Brooklyn and Harlem hospitals are compared. This is another indication of some possible localization of health effects of environmental conditions.

PROPOSED CONTINUED AND NEW WORK

Part of the third year will be devoted to completing the work still in progress, part to conducting research proposed in the original application but held in abeyance, and part to embarking on new avenues of research.

III. Continuing and Originally Proposed Projects

(6) Further work in identifying "time-space clusters" in acute health effects of environmental factors.

The tendency for asthmatics to have more attacks in the fall than in any other season has been reported by various investigators. Our own analyses (see item 4 above) show that these excess attacks in the fall are concentrated in short periods of time over and above the seasonal rise, and occur in different parts of a large metropolitan area simultaneously. These results serve as a means of screening for possible environmental agents. Only those that conform to the observed patterns, in this case only those agents concentrated within a few days time and acting simultaneously in both Brooklyn and Harlem, are plausible candidates as causal factors. Further work, to be performed by Dr. Wallenstein, is necessary to develop better statistical techniques to identify time and space clusters in disease that could serve as a screen for causal environmental factors.

(7) Further work on air pollution patterns in New York City

A "4-dimensional" grid of air pollution levels over New York City both in terms of height above street level and distance away from the inner-city center will be prepared. A computer program for generating such a profile will be developed under the direction of Drs. Goldstein, Landovitz and Fleiss.

(8) Space clustering

In the paper "Asthma Alley: A Space Clustering study of Asthma in Brooklyn, New York City" (accepted for publication, J.Asthma Res.), Dr. Goldstein has described a study of spatial clustering of asthma in the Bedford-Stuyvesant area of Brooklyn: one section of this area has a reputation, both among clinicians and lay persons, of having an unusually high rate of asthma attacks. In her study she demonstrated that when the asthma rates in the individual (small) health areas of Bedford-Stuyvesant were examined, those health areas having a higher than average rate (1) tended to form a single contiguous belt and (2) this contiguous belt coincided with the section called "asthma alley". It was intuitively obvious that there was only a very small probability that the health areas of high asthma prevalence could have formed a contiguous belt by pure chance, but methods of estimating this probability empirically have yet to be developed. This work will be done by Drs. Goldstein, Wallenstein and Fleiss.

IV. New Projects

(9) Variations in cancer mortality

The geographical distributions of cancer death rates are reported in the Atlas of Cancer Mortality for U.S. Counties, 1950-1969 (DHEW Publication No. NIH 75-780 by T.J. Mason, F.W. McKay, R. Hoover, W.J. Blot and J.F. Fraumeni). The Atlas contains maps of mortality rates for 16 common types of cancer on a county-by-county basis and of 19 other sites, for which a small number of deaths occurred during 1950-1969, by state economic area. Preliminary analyses of the above data point to various

possible environmental causal factors in the geographical distributions of cancer. National Cancer Institute scientists believe, for example, that population exposure to various industrial pollutants may be responsible for the marked geographic concentration of bladder cancers in the East (in New Jersey in particular). Melanoma rates vary with exposure to sunlight. Breast cancer and colon cancer show a similar geographic pattern, thus suggesting common environmental causal factors. High lung cancer rates have also been associated with counties containing smelting operations. These above average rates of lung cancer have been found both in men and in women, suggesting the possibility that the air pollution resulting from these industries affects the community surrounding these industries, in addition to the workers employed by them.

The above examples illustrate only a few of the clues to causal factors in environmental carcinogenesis that require further investigation. Very little beyond the comparison of age, sex and race specific cancer death rates by geographic areas has so far been published on the problem of developing formal criteria for the statistical significance of such clusters. Drs. Goldstein and Fleiss plan to search for testable hypotheses about possible causal factors, and to develop statistical tests for evaluating the tenability of these hypotheses.

(10) More on clustering

Drs. Goldstein, Landovitz and Wallenstein have begun preliminary analyses of two approaches to the problem of assessing the probability that a contiguous cluster of geographical subdivisions of high disease incidence or prevalence can have arisen by chance.

The first approach is based on random modeling. The contiguity of all geographical subdivisions of a large geographical area can be specified by a matrix. If a set of these subdivisions is found to have a high disease prevalence, the probability that by pure chance a smaller subset of this set is contiguous can be estimated by a random sampling procedure repeated a large number of times. The sampling process, which can be expressed formally in terms of matrix multiplication, can be carried out by computer.

The second approach requires that the subdivisions form a regular two-dimensional lattice. Contiguity is specified by the number of high disease incidence subdivisions that are, on the average, contiguous to a given high disease incidence subdivision. When the problem is stated in this form it can be shown to be equivalent to a well-known problem in statistical physics, that of the two-dimensional Ising model, to which exact solutions and approximate solutions of high computational accuracy are known.

It is our intention to explore both these approaches simultaneously, in a complementary way, as a means of estimating the a priori probability of clusters.

The second statistical problem that arises in studies of this type can be stated as follows. Suppose a certain geographical cluster of disease is established as unlikely to occur by chance, and therefore likely to reflect some environmental cause. Suppose further that some hypothesized causal factor -- say industrial pollution -- is to be evaluated. It would, of course, not be surprising if the geographical subdivisions of high industrial pollution are found to form a cluster also. This pollution cluster

may be found to overlap to some degree with the disease cluster. How much overlap is likely to occur by pure chance? This question is one that we will attempt to answer if our methods on the first type of question-- -- the statistical significance of an observed disease cluster -- are found to be effective.

(11) Time trends

To return to the problem of cancer mortality, suppose, for example, that the rates of cancer of the bladder rose in a certain county in the period 1965-1975 after a decade of stable rates. We would then most sensibly look for a change in environmental exposure somewhere in the 1940's to satisfy a latency period of approximately 20 years. Other carcinogenic theories would require other hypotheses. Statistical criteria for stability of death rate data over time need to be established, as well as for stability in environmental factors over time. Drs. Goldstein and Fleiss will assume responsibility for this work.

PAPERS WRITTEN OR IN PREPARATION

Cuzick, J. Asymptotic normality of chi-square in $m \times n$ tables with n large and small cell expectations. Results presented at spring statistics meetings, Chapel Hill, 1977.

Goldstein, I.F. Scientific base for sulfur oxide air pollution standard.

To appear in Proceedings of Sulfur Oxide Conference, John Carroll University.

Goldstein, I.F. Exposure data for the study of acute health effects of air pollution. Submitted to Archives of Environmental Health.

Goldstein, I.F., Goldstein, M., Fleiss, J.L. and Landovitz, L. Methodological problems arising from the choice of an independent variable in linear regression, with application to an air pollution epidemiological study. To be submitted to American Journal of Epidemiology.

Goldstein, I.F., Goldstein, M. and Landovitz, L. The relationship of air pollution to mortality: A critique. Published as a letter in the Journal of Occupational Medicine, 19:375, 1977

Goldstein, I.F., Landovitz, L., and Fleiss, J.L. Air pollution patterns in New York City, III. Spatial distributions of air pollutants. Paper in preparation.

Goldstein, I.F., Fleiss, J.L., and Landovitz, L. Air pollution patterns in New York City, IV. Principal components analyses. Paper in preparation.

Goldstein, I.F. and Rausch, L.E. Time series analysis of morbidity data for assessment of acute environmental health effects. Submitted to Environmental Research. An earlier version was presented at the 1977 meeting of the International Epidemiological Association, San Juan.

Stanford University1977 Progress Report and Plans for 1978

During the second year of the Study both its educational and scientific goals were advanced at the Department of Statistics, Stanford University.

Those participating were:

STANFORD UNIVERSITYStatistics Department

Persi Diaconis
 Ray E. Faith
 Yashaswini D. Mittal
 Lincoln E. Moses
 Ingram Olkin
 Thomas W. Sager
 Paul Switzer
 Alice Whittemore

Graduate Research Assistants

Keaven Anderson**
 Michael Chernick (Diaconis)*
 Delores Conway (Olkin)
 Richard DeVeaux (Diaconis)
 Chen-Hui Foo (Switzer)
 Cyril Grivet (Switzer)
 Bonnie Hole (Diaconis)
 Jack Homer*** (Sager)
 Elizabeth Kelly**
 Edward Korn (Switzer and Whittemore)
 Anthony Y. C. Kuk**
 William McCormick (Mittal)
 Alvaro Munoz (Mittal)
 Jeannette Prietsch (Sager)
 Javier Rajo**
 Ruth Sheshinski (Faith)
 Neil Willits (Olkin)
 Suzanna Wong (Moses)
 Yih Yuh**

Visitors

David Berengut (Harvard University)
 Ross Leadbetter (University of North Carolina)
 Donald Pierce (Oregon State University)
 Dag Tjøstheim (Norwegian Seismic Array, Norway)

*()'s indicate faculty member with whom graduate research assistant is working or by whom sponsored.

**Indicates first year graduate student - advisor to be assigned.

***Undergraduate student.

Seminar sessions, involving participating faculty and students as well as interested individuals from the Bay Area Air Pollution Control District (BAAPCD), California Air Resources Board (CARB), and the California Department of Health, were convened at regular intervals. These seminars covered a broad range of topics as shown in the following list; they served the useful purpose of information exchange with others primarily in the California area who have a continuing concern for environmental health problems.

SIMS SEMINARS

Speaker: DICK DEVEAUX (Stanford)*
Date: December 3, 1976 (too late to be listed in 1976 report)
Topic: "CORRELATION OF AIR POLLUTION VARIABLES"

Speaker: PROFESSOR THOMAS W SAGER (Stanford)*
Date: January 12, 1977
Topic: "SIMS STATISTICS AND AIR POLLUTION SEMINAR"

Speaker: PROFESSORS PAUL SWITZER, RAY FAITH, PERSI DIACONIS,
TOM SAGER (Stanford)*
Date: January 19, 1977
Topic: A series of short talks by members of the SIMS group.

Speaker: MIKE KIM (Transportation Planning Engineer LIRAQ)
Date: January 26, 1977
Topic: "DEVELOPMENT OF SOURCE EMISSIONS INVENTORY FOR LIRAQ"

Speaker: PROFESSOR RAY FAITH (Stanford)*
Date: February 2, 1977
Topic: "MORE ON KRIGING"

Speaker: PROFESSOR DAG TJUSTHEIM (Norsar Institute)*
Date: February 9, 1977
Topic: "A NEW MEASURE OF ASSOCIATION FOR SPATIAL VARIABLES"

Speaker: PROFESSOR PERSI DIACONIS (Stanford)*
Date: February 16, 1977
Topic: "A CLASSICAL MATHEMATICAL STATISTICS PROBLEM WHICH ARISES NATURALLY IN AIR POLLUTION"

Speaker: ED KORN and ALICE WHITEMORE (Stanford)*
Date: February 23, 1977
Topic: "PROBLEMS RELATED TO ASTMA AND AIR POLLUTION"

* Members of the SIMS Study

Speaker: CHARLES UNGER (Planning Division, Air Resources Board)
Date: March 4, 1977
Topic: "THE RELATIONSHIP BETWEEN THE STATISTICAL PARAMETERS OF POLLUTANT CONCENTRATIONS AND METEOROLOGICAL PARAMETERS"

Speaker: BONNIE HOLE AND TOM SAGER (Stanford)*
Date: March 9, 1977
Topic: "PERSONAL AND PROFESSIONAL REACTIONS TO LIRAQ"

Speaker: CYRIL GRIVET (Stanford)*
Date: April 8, 1977
Topic: "OXIDANT AND CO IN THE BAY AREA"

Speaker: ED KORN (Stanford)*
Date: April 15, 1977
Topic: "CONTINGENCY TABLES WITH CLASSIFICATION ERRORS: A STATISTICAL PROBLEM ARISING IN ANALYSING THE RELATIONSHIP BETWEEN AIR QUALITY AND HEALTH"

Speaker: MICHAEL CHERNICK (Stanford)*
Date: April 22, 1977
Topic: "AN APPROACH TO DETECT CHANGES IN A TIME SERIES MODEL FOR MONTHLY OZONE LEVELS"

Speaker: TOM SAGER AND OTHER SIMS FACULTY (Stanford)*
Date: April 29, 1977
Topic: A Round Table Discussion

Speaker: NEIL WILLITS (Stanford)*
Date: May 6, 1977
Topic: "THE BRUTE FORCE METHOD OF AIR POLLUTION MODELING"

Speaker: LEWIS H. ROBINSON (Bay Area Air Pollution Control District)
Date: May 13, 1977
Topic: "PRINCIPAL COMPONENTS, POLLUTION INDICES, AND METEOROLOGICAL FIELDS"

Speaker: CHEN-HUI FOO (Stanford)*
Date: May 20, 1977
Topic: "THE IMPROVEMENT OF AIR POLLUTION STUDY BY USING WEIGHTED LEAST SQUARE AND TRIMMED MEAN METHODS"

Speaker: H. JEAN THIEBAUX (Department of Mathematics, Dalhousie University, Halifax, Nova Scotia)
Date: May 23, 1977
Topic: "STATISTICAL STRUCTURE AND ATMOSPHERIC DYNAMICS"

*Members of the SIMS Study

Speaker: SUZANNA WONG AND JEANNETTE PRIETSCH (Stanford)*
Date: June 3, 1977
Topic: "AN OVERVIEW OF THE 1977 BERKELEY INTER-AGENCY SYMPOSIUM ON AIR MONITORING QUALITY ASSURANCE"

Speaker: PROFESSOR PAUL SWITZER (Stanford)*
Date: June 29, 1977
Topic: "A REVIEW OF SOME RECENT SIMS RESEARCH ACTIVITIES AT STANFORD"

Speaker: PAUL SWITZER (Stanford)*
Date: July 6, 1977
Topic: "ESTIMATING SPATIAL DISTRIBUTION FUNCTIONS"

Speaker: JACK HOMER (Stanford)*
Date: July 13, 1977
Topic: "NORMAL VS LOGNORMAL DISTRIBUTIONS FOR POLLUTANT CONCENTRATIONS" - PART 1

Speaker: JACK HOMER (Stanford)*
Date: July 20, 1977
Topic: "NORMAL VS LOGNORMAL DISTRIBUTIONS FOR POLLUTANT CONCENTRATIONS" - PART 2

Speaker: M. R. LEADBETTER (Stanford and University of North Carolina)*
Date: July 27, 1977
Topic: "A COMPARISON OF EXTREMAL THEORY FOR CONTINUOUS AND SAMPLED STOCHASTIC DATA"

Speaker: RAY FAITH (Stanford)*
Date: August 3, 1977
Topic: "OPTIMAL INTERPOLATION: ADVANCES AND RETREATS"

Speaker: RICHARD KOPPEL (Dept. of Operations Research, Stanford)
Date: August 5, 1977
Topic: "BOX-JENKINS MODELLING OF AIR QUALITY"

Speaker: DAVID AHN (Stanford)*
Date: August 10, 1977
Topic: "DOES RADIATION IN OUR ENVIRONMENT AFFECT OUR HEALTH?"

Speaker: FRANK LUDWIG (Senior Research Meteorologist, Atmospheric Sciences Laboratory, SRI International)
Date: October 5, 1977
Topic: "METEOROLOGICAL ASPECTS OF OZONE DISTRIBUTIONS IN THE UNITED STATES"

Speaker: DONALD HARDY and JOHN WALTON (Lawrence Livermore Laboratory, University of California)
Date: October 12, 1977
Topic: "NEW APPLICATIONS OF PRINCIPAL COMPONENTS ANALYSIS IN APPLIED METEOROLOGY"

Speaker: MICHAEL CHERNICK (Stanford)*
Date: October 19, 1977
Topic: "LIMIT DISTRIBUTIONS FOR MAXIMA OF STATIONARY SEQUENCES"

Speaker: CYRIL GRIVET (Stanford)*
Date: October 26, 1977
Topic: "MODELS FOR THE MAXIMA OF POLLUTION CONCENTRATION"

Speaker: MICHAEL AXELROD (Lawrence Livermore Laboratory)
Date: November 2, 1977
Topic: "ASSESSING THE AIR QUALITY IMPACT OF GEOTHERMAL ENERGY"

Speaker: J. R. MARTINEZ, Senior Research Engineer, Atmospheric Sciences
Laboratory, Stanford Research Institute
Date: November 9, 1977
Topic: "MODELING PHOTOCHEMICAL OXIDANT BY A MARKOV CHAIN"

Speaker: Discussion led by Donald A. Pierce*
Date: November 16, 1977
Topic: "APPROACHES TO ANALYSIS OF CHES ASTHMA DATA"

Speaker: Discussion led by Alice Whittemore*
Date: November 23, 1977
Topic: APPROACHES TO ANALYSIS OF CHES ASTHMA DATA

* Members of the SIMS Study

Colloquia, Conferences, and Meetings

Invited papers were presented at conferences and professional society meetings by faculty members of the study relating to their research on air pollution problems. These are listed below. In addition, members of the study were invited participants at the annual conference on Quality Assurance for Air Pollution Monitoring Data held in Berkeley and sponsored by the California Department of Health (Faith, Mittal, Sager, Switzer in 1976 and Prietsch, Wong, Faith, Foo in 1977). In June, 1977, the Western Regional Meeting of the Institute of Mathematical Statistics and the Biometric Society co-sponsored a session on Statistics for Environmental Problems with Switzer as chairman and Mittal as an invited speaker. Two members of the group (Sager, Hole) participated in a workshop at the Lawrence Livermore Laboratory on the application of the LIRAQ air pollution simulation model in May, 1977. Three group members (Moses, Whittemore, Korn) participated in a weekend study session organized by John Goldsmith of the California Department of Health in May, 1977. In December, 1977, Switzer will present an invited paper at the session on Spatial Statistics at the biannual meeting of the International Statistical Institute in Delhi. Lincoln Moses attended the Third ERDA Statistical Symposium, held in Richland, Washington, and participated in the discussion of how best to establish an environmental monitoring program at the Oil Shoal project now under development in Utah.

(1) Yash Mittal

"Use of averaging maxima in air pollution data"
Fourth Symposium on Statistics and the Environment,
Washington, May, 1976. (not listed in 1976 report)

(2) Yash Mittal

"The joint distribution of the maxima of two air pollution time series" (with J. M. Cuzick, Columbia University)
Biometric Society and IMS joint meeting,
Stanford, June, 1977.

Colloquia, Conferences, and Meetings (cont.)

(3) Thomas W. Sager

"A new approach to regression in certain time/space series problems"

SIMS Research Application Conference on Time Series and Ecology
Alta, Utah

June 27-July 2, 1977

"Statistical issues in air pollution assessment"

Air Pollution Seminar, Department of Meteorology
California State University, San Jose

November 30, 1977

(4) Paul Switzer

"Distribution of air pollutants in the San Francisco Bay Area"
Center for Geostatistics, Fontainebleau, France

September 20, 1977

"Estimation of spatial distributions from point sources
with application to air pollution measurement"

International Statistical Institute, Delhi, India
December 6, 1977

(5) Alice S. Whittemore

"Mathematical models of carcinogenesis: epidemiological
implications"

University of California, Los Angeles
February, 1977

University of California, Riverside
March, 1977

California State University, Hayward
March, 1977

M. D. Anderson Hospital and Tumor Institute
Houston, Texas
December, 1977

Data Acquisition

The following steps were taken with respect to data acquisition:

- (1) Pierce went to Los Angeles in early October and conferred with Roger Detels, Professor of Epidemiology at UCLA School of Public Health, concerning the possibility of mobile station data for comparison with fixed station data.
- (2) Whittemore and Thomsen (SIMS) conferred in December with Warren Winkelstein, Dean and Professor of Epidemiology, Susan Sacks, and Steven Selvin at the School of Public Health, University of California, Berkeley, and with Craig Hallowell, Deane Merrill, Donald Austin, Simcha Knif, and Ruth Levine from Computer Science at Lawrence Berkeley Laboratory concerning secondary health and air pollution data (known as "Populations at Risk") on a county-by-county basis across the United States.
- (3) Thomsen conferred with William Nelson of EPA and with Clifford H. Patrick at ERDA, turning up several leads, which remain to be followed up.
- (4) While on leave at the Center for Geostatistics, Paul Switzer has assembled a file of papers on air pollution research being done in France.
- (5) Switzer has arranged for the acquisition of data from a dense monitoring network in Venice, Italy, in conjunction with forthcoming visit by Paolo Zannetti of the Venice IBM Scientific Center.

Research Activities (1977 and proposed for 1978)

In reviewing research activities for the second year and those proposed for 1978, wherever appropriate reference is made through the use of relevant section numbers to the original proposal, "Proposal for SIMS Three-Year Study on Statistics and Environmental Factors in Health," dated February, 1976. Many of the research lines initiated or in progress during the past year will be continued in 1978.

Olkin studied methods for calculating or approximating the probability that at least one of a series of correlated pollutants exceeds its threshold. He has proposed several procedures for this problem and other closely-related problems for both normally transformed data and untransformed data.

He has also been studying the possible application of modern cartographic methods to the display of air pollution fields to better understand the variability of exposure. In addition, he is exploring the choice of models for joint distributions of several pollutants or several monitoring stations considered simultaneously.

Tjøstheim (visitor) investigated the problem of correlating the concentrations of two different pollutants measured at the same station network. He proposed a new correlation measure based on ranks which explicitly also employs station-to-station distances.

Mittal's theoretical research (§2.1.3) involves possible application of maxima of partial sums from stationary Gaussian sequences to air pollution by attempting to devise an extreme value theory of such sums. Extreme value theory is often applied to air pollution data; however, it depends on what distribution the observations come from. Problems arise because of the difficulty of fitting any fixed distribution to the data. To rectify this, it is suggested that extreme value theory should be applied to

suitably normalized data sequences rather than the original data sequences. Mittal worked at the theoretical problem of devising an extreme value theory for the normalized sequence. She plans to pursue this problem in 1978 and apply it to real data. Currently she is supervising Munoz (graduate student), who is analyzing Bay Area oxidants.

Switzer, on leave at the Center for Geostatistics, Fontainebleau, France, has been working on how to take account of weather variables in the study of pollution variables, and his planned work lies in the area (§2.1.3, 2.3.1). The notion of random regression functions may provide a powerful alternative analytic tool to the usual linear regression theory which often seems inadequate.

He will also be concerned with the consequences of applying statistical methods derived from stationary probability assumptions to air quality data which have rather obvious spatial trends and time periodicities.

Faith (§2.1.1) has derived some further results concerning the optimal interpolation of a stationary stochastic function on the basis of current and past values of the function at specific locations. These extend the previous results by incorporating the time element and by permitting the spatial mean function to be relatively general. He developed computer programs to plot isopleths of oxidant readings and worked jointly with Ruth Sheshinski to provide data based examples of the effects of misspecification of the mean function. He applied cross-validation criteria to the estimation of the spatial correlation function and to the assessment of the accuracy of interpolation. Faith is writing up the theoretical results mentioned above. In addition to exploring further the assessment of interpolation methods via cross validation he intends to study in the

coming year the degree to which the pollution measured at a monitoring station is representative of its value in the local vicinity by applying similar methods to measurements taken at points relatively close to each other.

Sager (§2.1.1, 2.1.4, 2.2.0) has been working on extending his method (Technical Report #1) for analyzing distributions of air pollutants and their effects on health; reference here is to the effect on health of spatial variation of pollution. Technical Report #11 analyzes the effect of temporal variation of pollution. Both methods are based on the idea of comparing high and low isopleths of the pollutant distribution to corresponding high and low health isopleths. Of particular interest in Technical Report #11 is the nonparametric (or model-free) time-series analysis. Jack Homer has finished work under Sager's direction on methods of discriminating between normal and lognormal models for air pollution concentrations in the San Francisco Bay Area. This work will appear as a working paper. For 1978, further development of isopleth methods may be expected. More inferential tools will add to the practical utility of the method. Additionally, a start will be made on the problem of optimal siting strategy for air monitoring.

Homer (undergraduate student) (§2.3.1) has done a number of original calculations for the fitting of normal and lognormal distribution models to air quality time series.

Pierce has developed a method for combining several samples (different stations, or different time periods) in testing goodness of fit of air pollution variables to the lognormal distribution. He has also begun to

develop an approach to analyzing the EPA-CHESS asthma panel data by applying Cox Likelihood methods to the point process defined by beginning- (or beginning- and ending-) points of asthma attacks.

Whittemore, together with SIMS graduate student Korn, is analyzing some of the asthma panel studies conducted by EPA-CHESS. Also, in collaboration with J. B. Keller, she is developing a mathematical theory for the biological mechanisms of bronchial asthma. The predictions of the theory will be compared to the EPA-CHESS data. In addition, she edited the SIAM/SIMS Volume, Environmental Health: Quantitative Methods (Proceedings of SIMS 1976 Conference on Environmental Health) which was published in 1977.

Wong, working under the direction of Moses, is nearing completion of her dissertation on extrapolation of bioassay data to extremely low doses.

Grivet (graduate student) analyzed air quality and weather data simultaneously in order to assess the degree to which certain weather variables may be used to predict pollutant concentrations. A considerable amount of relevant data was made available for this problem by the BAAPCD. Leadbetter (visitor), from the theoretical mathematical point of view, and Grivet, using data analytic techniques, both looked at the problem of assessing the difference between maximum concentrations of a continuously-recorded measurement vis-a-vis the maximum determined from periodic observations.

In 1978 Grivet will be working on joint distributions of the concentrations of several pollutants as well as making a critical examination of Larsen's empirical models for distributions of maximum concentrations.

Technical Reports summarize lines of research that are more fully developed than what is represented in the Working Papers. Both types published to date plus related external publications are listed below:

SIMS Technical Reports

- #6 Dag Tjøstheim (in process)
"On the Problem of Estimating the Autocorrelation Function of Spatio-Temporal Variables"
- #7 Dag Tjøstheim (March 17, 1977)
"A New Measure of Association for Spatial Variables"
- #8 R. Bohrer, W. Chow, R. Faith, V.M. Joshi, C.F. Wu (March 22, 1977)
"Multiple Three-Decision Rules for 2^k -Factorial Simple Effects"
- #9 Paul Switzer (July 1, 1977)
"Estimation of Spatial Distributions from Point Sources with Application to Air Pollution Measurement"
- #10 M. R. Leadbetter (August 1, 1977)
"On Extreme Values of Sampled and Continuous Stochastic Data"
- #11 Thomas W. Sager (October 14, 1977)
"A New Approach to Regression in Certain Time/Space Series Problems"
- #12 Alice S. Whittemore (December 6, 1977)
"Collapsibility of Multidimensional Contingency Tables to Tables of Smaller Dimension"
- #13 Alice S. Whittemore (December 9, 1977)
"The Age Distribution of Human Cancer for Carcinogenic Exposures of Varying Intensity"
- #14 Michael Chernick (in process)
"A Limit Theorem for the Maximum of an Exponential Autoregressive Process"

Working Papers

- #6 Chen-Hui Foo (September 1977)
"Application of Some Robust Statistical Procedures to Air Pollution Data in the San Francisco Bay Area"
- #7 Jack B. Homer (in process)
"Techniques for Discriminating Between Normal and Lognormal Populations with Special Application to Air Pollution Data"

External Publications (papers published in journals, conference proceedings, etc.)

"Relating Spatial Distributions of Pollutants to Health Effects,"
Thomas W. Sager, Proc. 9th International Biometrics Conference, Vol. II,
pp. 35-58, The Biometrics Society, 1976.

"A New Approach to Regression in Certain Time/Space Series Problems,"
Thomas W. Sager, to appear in Proceedings of SIMS Research Application Conference in Time-Series and Ecology, June 27-July 1, 1977. Alta, Utah.

External Publications (cont.)

"Maxima of Partial Samples in Gaussian Sequences," Yashaswini Mittal, (accepted for publication in The Annals of Probability).

"The Age Distribution of Human Cancer Corresponding to Carcinogenic Exposures of Varying Intensity," Alice S. Whittemore, Am. J. Epid. 106: 418-532, 1977.

"Collapsibility of Multidimensional Contingency Tables to Tables of Smaller Dimension," Alice S. Whittemore. (submitted to J. Royal Statist. Soc., Ser. B.)